DOCUMENT RESUME

ED 319 540 PS 018 849

AUTHOR Ellwein, Mary Catherine; And Others
TITLE Using Readiness Tests To Route Kindergarten Students: The Snarled Intersection of Psychometrics, Policy, and Practice.
PUB DATE 90
NOTE 46p.; Paper presented at the American Educational Research Association Annual Conference (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Age Differences; Educational Practices; *Ethnicity; *Kindergarten; Primary Education; School Districts; School Policy; *Screening Tests; Sex Differences; Socioeconomic Status; *Student Placement; Test Bias; *Test Reliability; Test Use

ABSTRACT

The validity of inferences drawn from readiness tests used to place children in a regular or a 2-year kindergarten program was investigated. Data from nine school districts in Virginia were used in the study. The study explored technical characteristics of four commonly used readiness tests: the Brigance Kindergarten and First Grade Screen, the Developmental Indicators for the Assessment of Learning-Revised (DIALR), the Daberon Screening for School Readiness, and the Missouri Kindergarten Inventory of Developmental Skills (KIDS). Findings indicated that male, minority, low socioeconomic status, and young children scored consistently lower than others on all four tests. Three of the tests were fairly reliable, but the KIDS was highly inconsistent for minorities and those under 5 years of age. For these children, true scores could be nearly 20 percent higher or lower than those obtained. Although the Brigance, Daberon, and DIALR scores were more consistent indicators than anticipated, they fell short of standards which recommend reliability coefficients of .90 or higher for placement decisions. No test was an impressive predictor of future test performance. Findings are interpreted in the context of specific selection and placement policies. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319540

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Using readiness tests to route kindergarten students: The snarled intersection of psychometrics, policy, and practice

Mary Catherine Ellwein
University of Virginia

Daniel J. Walsh
University of Illinois

Gerald M. Eads, II
Virginia Department of Education

Alexandra Miller
Virginia Department of Education

PS 018849

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Daniel J. Walsh
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

We examine the validity of inferences drawn from readiness tests when placing children in the regular or a two-year kindergarten program. Using data from nine Virginia school districts, we studied the technical characteristics of four commonly-employed readiness tests. Males, minorities, low SES, and young children scored consistently lower on all four tests. Three of the four tests were fairly reliable, but the fourth was highly inconsistent for minorities and those younger than five. None of the four tests were impressive predictors of future test performance. The results are interpreted within specific selection and placement policies.

Using readiness tests to route kindergarten students: The snarled intersection of psychometrics, policy, and practice

Contemporary kindergarten classrooms are no longer the gardens in which the intellectual, social, and personal development of young children are gently and equitably tended. In their place, boot camps have been established by which students are inducted and instructed in a narrow academic curriculum to prepare them for the demands of first grade and future schooling. Just as the armed services screen their charges for physical and mental competence prior to sending them to boot camp and their tour of duty, so too are the schools requiring young children to demonstrate that they will be able to endure the rigors of the kindergarten curriculum, basic to their schooling career.

Despite calls to the contrary (NAEYC, 1988; NAECSSDE, 1987), legislation in 16 states and more than half the districts in another seven states compel schools to conduct kindergarten entrance screening (Cannella & Reiff, 1989). Tests are used to determine who may enter kindergarten as well as identify those in need of remediation or those who qualify for advanced placement.

Activities in one of these states illustrate the legislated role of testing in determining who may enter a regular kindergarten classroom and who may not. Until 1986, Virginia schools admitted all children as long as they were five years old

on or before December 31 of the academic year. In the following three years, the cutoff was rolled back in monthly increments until only those born before October 1 were automatically eligible to come to kindergarten. Those born after the new date(s) and December 31 could be enrolled only after taking one of five state-approved tests. School districts retained sole power in determining test performance standards and any other criteria by which these children could be enrolled.

In response to the compulsory screening, voluntary testing escalated. More than one-third of Virginia divisions elected to test all age-eligible as well as petitioning children (Eads, Miller, Ellwein & Walsh, 1990). Although districts could not deny admission to the former, many used test results for placement decisions. Children who did not perform well on the tests were considered "unready" and were subsequently placed in junior kindergarten, an extra-year placement to be followed by a year in regular kindergarten. In the districts we studied, anywhere from 11% to 36% of those tested were, in the vernacular, pre-flunked. That is, on the basis of screening conducted during the summer or fall preceding the kindergarten year, these children were routed through a junior kindergarten-regular kindergarten track.[1]

Problem

The central role of testing in the policies and practices of kindergarten entry prompts several compelling questions. In this paper, one serves as the organizing query: How valid are the inferences drawn from readiness tests when charting the initial routes of children's schooling careers? Specifically, we address five questions:

1) What is assessed on tests employed for this purpose?

2) How does test performance vary according to the gender, ethnicity, and SES of those examined?

3) To what extent can test performance be attributed to differences among children rather than measurement error?

4) To what degree does test perfcrmance correspond with future school measures?

5) When school policies or practices allow for the use of tests when making kindergarten placement decisions, what consequences may ensue?

Perspective

In addition to the research summarized in the following paragraphs, our work is informed by Cronbach's conceptualization of test validation (1980a, 1980b, 1984). As reflected in the organizing query, we do not speak of the validity of these tests in a nomothetic sense. Rather, we seek to examine the "soundness and relevance of a proposed interpretation of scores" given the

contexts in which the tests are used (Cronbach, 1984, p. 125). Second, we acknowledge that empirical efforts such as those reported here are but one line of reasoning in arguing for validity; judgment and values must also be part of the discourse (House, 1980). Third, we concur with Cronbach when he stated that the "validity of an interpretation cannot be established [emphasis in the original] by a research monograph . . . . the aim for the report is to advance sensible discussion" (1980b, p.107). Moreover, such discussion would be incomplete without deliberations over the soundness of the selection procedures of which these tests are a part.

Related research

Two distinct lines of research inform the present study. Traditional educational psychology and psychometric forums have regularly published work that examined the criterion-related validity of an array of individual readiness tests and batteries (Butler, Marsh, Sheppard, & Sheppard, 1982; Gordon, 1988; Harrison & Naglieri, 1981; May, 1986; Popovics, 1982; Rubin, Barlow, Dorle, & Rosen, 1978; Schmidt & Perino, 1985; Stoner & Purcell, 1985). Fifty-eight such studies formed the basis for a meta-analysis conducted by Horn and Packard (1985), the purpose of which was to identify variables with the greatest predictive validity. They found that achievement in first grade was best forecasted by ratings of distractibility and "internalizing"

behavior problems (r=.71 and r=.56) followed by performance on written expressive language tasks (r=.61) and group IQ tests (r=.59)-- note that these predictors were studied in isolation, no multivariate analyses were conducted. Measures of gross motor development, an important indicator of future school success according to conventional wisdom, were among the weakest predictors. Although most of the individual studies are well-suited for inclusion in technical manuals and the meta-analysis points to generic predictors, they offer little or no insight on the use or interpretations of test scores when making placement decisions.

The one exception to this conclusion points to the second, more relevant, line of research which addresses the link between assessing readiness and placement decisions. Because the Gesell School Readiness Test has been promoted by its authors as a placement test and its use was widespread, it has been the focus of the greatest amount of study in recent years. Drawing on technical studies and reviews of the Gesell (Bradley, 1985; Kaufman & Kaufman, 1972; Kaufman, 1975; Naglieri, 1985; Popovics, 1982; Shepard & Smith, 1985; Wood, Powell, and Knight, 1984), researchers have concluded that the Gesell lacks discriminate validity from IQ tests and its use can result in the misidentification of 50-60% of the children tested and placed in extra-year programs (Graue & Shepard, 1989; Shepard & Smith, 1985).

The likelihood of misidentification with the Gesell exacerbates problems associated with the effectiveness of extra-year programs, notably kindergarten retention and transition first-grade placements. Shepard and Smith (1987) found no differences between children retained in kindergarten and their matched promoted peers on standardized math scores, teacher ratings of reading and math achievement, learner self-concept, social maturity, and attention span. A one-month difference (in favor of the retained children) was evident on standardized reading scores. In her review of research evaluating transition first grade classrooms, Gredler (1984) observed that in four of the five studies children did not benefit from their transition-room placement when compared to their promoted peers. To date, no studies have been published on the effects of immediate entry and participation in a two-year kindergarten track.

Concerns about the utility of the Gesell and effects of extra-year programs have led to the generalization that there exist no readiness tests with adequate reliability or validity to warrant special placement decisions (Shepard & Smith, 1988; Meisels, 1989). Although we are inclined to agree, we believe it best to extend the empirical basis for the conclusion by subjecting other readiness tests to a comparable level of scrutiny as paid to the Gesell. Moreover, if experiences in the state of Virginia are any indicator of broader trends, the use of the Gesell has declined dramatically while the use of other

readiness tests has increased. Anecdotal evidence from school districts and the State Department of Education attribute the decline not only to effective communication of research findings but to the considerable investment of dollars and time in the use of the Gesell.

In the following paper, we first address the technical characteristics of four popular readiness tests: the Brigance Kindergarten and First Grade Screen, the Developmental Indicators for the Assessment of Learning-Revised, the Daberon Screening for School Readiness, and the Missouri Kindergarten Inventory of Developmental Skills. Using data obtained from nine Virginia school districts, we examine the extent to which screening test performance varied according to child characteristics including gender, race, social class, and age. Next, we report on the extent to which each test produced reliable or consistent scores. Criterion-related validity is investigated by analyzing the correspondence between screening test scores and end-of-kindergarten as well as early first grade performance. Finally, we situate the technical concerns within the broader policy and practice issues. That is, we discuss the educational and social implications of using test results to determine who enters regular kindergarten and who sets foot in a junior kindergarten. Before pursuing these issues, however, we digress briefly to describe the four tests as well as the methods used to study them.

Tests Used for Kindergarten Placement

In the following paragraphs, we describe the purpose(s) and characteristics of each test. Table 1 summarizes the skills assessed across the four tests and Table 2 shows the type of self-critical evidence (e.g., reliability and validity studies) provided in the test manuals.

Table 1 about here

Brigance Kindergarten & First Grade Screen

The Brigance was designed to provide an easy and quick assessment of a student's development in skill areas that have "the greatest predictive validity for success in kindergarten and first grade" (Brigance, 1982, p. iii). However, the technical manual offers no evidence in support of this claim. According to the manual, the Brigance may be used for one or more reasons including referral, placement, and instructional planning. Individually administered in 10 to 20 minutes, the test covers 12 subtests at the kindergarten level and yields a possible total score of 100.

Daberon Screening for School Readiness

The Daberon was designed to "sample the knowledge and skills of children as they begin school" (Danzer, Gerber, & Lyons, 1982,

p. 6). According to the test authors, the Daberon is intended to predict school readiness; it should not be used for diagnostic evaluations since it does not "test a child to his limits" (p. 6). Individually administered in 20 to 40 minutes, the test yields two indices: 1) a total raw score which ranges from 0 to 122 and a Readiness Equivalency Age (REA) which is derived from the correlation between total score and chronological age.

Developmental Indicators for Learning-Revised

The DIALR was designed as a screening test to identify children in need of additional evaluation. The authors state that the DIALR should not be used for intervention, placement, or treatment although it could be used in the regular classroom to identify individual strengths and weaknesses as the first step in developing curriculum and planning instruction (Mardell-Czudonowski & Goldenberg, 1983). Individually administered in 20 to 30 minutes, the DIALR covers motor, conceptual, and language skills in children two to six years old. Scaled scores ranging from 0 to 4 are assigned to each task and summed within skill area. A total scaled score is the sum of scaled scores across the three equally-weighted skill areas (maximum points possible: 93).

Kindergarten Inventory of Developmental Skills

The KIDS is said to provide "a systematic method for obtaining information on [a] child's development during a time of rapid growth and learning" (Missouri Dept of Elementary & Secondary Education, 1981, p. 1). The authors treat the KIDS as a means of screening children for one or more reasons including identifying children in need of further evaluation; suggesting areas where parents can initiate learning activities; and assisting teachers in planning appropriate experiences for every child. The instrument can be administered to an individual child in 35 minutes or less. Both subtest and total scores are the sums of the correct number of responses or tasks completed. Subtest scores may be converted and reported as stanines, percentiles, and normal curve equivalents.

Table 2 about here

Method

Sample

School Districts. Among the 50 districts that screened all eligible and petitioning children, nine contributed data on all those screened for the 1986/87 academic year. Compared to the districts that did not participate in the study, the sample districts enrolled fewer children (Median=1,852; M=1,962; SD=987

vs Median=2,801; M=5,911; SD=8,875) and a slightly smaller proportion of students eligible for lunch subsidies (M=.35, SD=.12 vs M=.40, SD=.17). Table 3 reports district-level characteristics including enrollment, proportion of children eligible for subsidized lunch, type of kindergarten programs available, screening test employed, and application of cut-scores.

Table 3 about here

Table 3 also describes the characteristics of the children who were screened and their initial placements in each of the nine districts and Table 4 aggregates these numbers by screening test. With the exception of those assessed with the Daberon, a slight majority were boys. Minorities constituted 45% of the Brigance subsample, but accounted for proportionately less in that of the Daberon and KIDS.[1] There were no minority students in the district using the DIALR. With the exception of those screened with the KIDS, 30% or more of the students were eligible for free or reduced lunch. All screening test subsamples included students who were born before 1981. The remaining students were fairly evenly distributed across the first three

---

[1] Due to small numbers, Native Americans and Hispanics were grouped with blacks to constitute a minority comparison group. Asians were not included in this group as they are not typically considered an academically "disadvantaged" minority.

quarters of 1981 with relatively fewer born in the last three months of the year.

Table 4 about here

In 1986/87, 78% of the children were placed in regular kindergarten classrooms (n = 1,007), six of whom were repeating that grade. More than one-fifth were sent to junior kindergarten. Only four children were not enrolled in these schools; they were all December-born and according to the state policy, the schools were not obligated to admit them. In 1987/88, 75% were in regular first grade classrooms. Fifteen percent were in regular kindergarten, most having spent the prior year in junior kindergarten. Another 6% were either retained in kindergarten or placed in a transition first grade classroom. One percent were placed in special education and 3% of the 1987/88 placements were missing.

Criterion Measures

Exit kindergarten performance. The Metropolitan Readiness Test (MRT) is a measure of readiness for first grade (Nurss and McGauvran, 1986). Level two of the test (MRT2) examines auditory, visual, and language abilities in separate subtests, the scores of which are combined to produce a pre-reading composite. A fourth subtest is designed to measure quantitative

abilities. Although derived scores are available on the MRT2, pre-reading composite and quantitative subtest raw scores were used in the analyses.

Early first grade performance. The Cognitive Abilities Test, Level 1 (CogAT) was designed to "appraise the full range of individual differences in the development of general cognitive skills among children in grades K-3 (Thorndike & Hagen, 1987, p.4). In measurement circles, the CogAT is accepted as a "curriculum-free" test of learned abilities (Loyd, personal communication). According to state mandate, the CogAT is routinely administered to all students in the fall of their first grade year. Three general areas are assessed on the CogAT: verbal, quantitative, and nonverbal; for the purposes of this paper, we used a composite CogAT score derived from performance in these areas.[2] According to the test authors, the composite score reflects the level of general reasoning ability possessed by an individual.

Results

Screening test performance

For each screening test, we looked at the degree to which total scores differed according to gender, ethnicity, SES, and age cohort. Table 5 shows the means and standard deviations by test. Table 6 provides effect sizes to illustrate the magnitude of test score differences between groups (e.g. poor and non-poor)

and to compare these differences across measures, (Glass & Hopkins, 1984). The determination of statistical significance and explained variance was based on the use of analysis of variance (ANOVA) techniques for unbalanced designs. Although the ANOVA results inform the following section, in the interest of brevity we confine our remarks to the patterns observed across the four tests (see Table 6). We refer the interested reader to the appendix for the individual ANOVA tables and interpretations.

Table 5 about here

Differences in child characteristics were related to differences in readiness test performance. Although females outscored males on each test, the difference was significant only on the Brigance in which males scored one-third of a standard deviation lower than females (-.31 Δ). Minorities scored consistently lower than whites on three of the four tests (recall that no minorities were included in the DIALR sample). The discrepancy was greatest on the KIDS where whites scored over one standard deviation higher. Minorities scored -.63 Δ and -.13 Δ on the Brigance and Daberon, respectively. Poorer children scored lower than their non-poor counterparts on every test, although the difference was not statistically reliable on the KIDS. Children eligible for free or reduced lunch averaged a

standard deviation lower on the Daberon, followed by -.69 Δ and -.56 Δ on the Brigance and DIALR, respectively.

Table 6 about here

Considering only gender, ethnicity, and SES, explained variance ranged from 15% in the Brigance to 24% in the Daberon. When age was added to the analyses, explained variance increased appreciably, ranging from 26% on the Brigance to 52% on the KIDS (see appendix). Not surprisingly, older children scored consistently higher than their younger peers. Winter-born children scored anywhere from .81 to 1.39 Δ higher than those born in the last quarter of 1981. Differences decreased between spring- and fall- as well as summer- and fall-born children on each of the readiness tests.

Reliability of screening tests

Using item-level data, internal consistency estimates and standard errors of measurement (SEMs) were computed on the levels at which scores on each test were typically reported. For each test, Table 7 summarizes the reliability coefficients (Cronbach's coefficient alpha) by gender, ethnicity and age.[3] Table 7 also presents standard errors of measurement (SEM) as the statistic is a more meaningful indicator of measurement precision than the reliability coefficient (Hopkins & Stanley, 1981).

Table 7 about here

Coefficient alphas ranged from .81 on the Brigance to .89 on the DIALR. The Brigance was equally ccnsistent for males and females, minorities and whites, and younger and older examinees. The SEM for the entire group was 8% although whites and those younger than five when screened had a slightly lower SEM (7%). Similarly, the Daberon and DIALR were consistent across gender, ethnicity, and age at time of screening, but each was somewhat more precise than the Brigance (SEM = 5%). Finally, the KIDS appeared to be the most inconsistent among the four screening tests as the SEM for the entire sample was 14%. Greater error attended scores of minorities and those younger than five when screened. Standard errors for these two groups approached 20% whereas scores for whites and those older than five when screened were considerably smaller: 12% and 7%, respectively.

Predicting readiness for first grade

The sample districts had varying policies about end-of-K testing. All but the KIDS district (#72) administered some form of readiness test and seven used some form of the Metropolitan Readiness Test. Various combinations of junior and regular K students took the 1974 edition or one of two levels of the 1986 edition of the MRT. Due to this lack of uniformity, we limited

our analyses to those who were placed in kindergarten in 1986/87 and had scores on the 1986 edition of the MRT2. As a result the MRT2 analyses were based on data from five districts (#11, #54, #90, #99, and #207). Although the decision to exclude MRT2 scores of junior K students incurred some range restriction in those districts with dual-track programs, the decision allowed us to avoid the confounding of treatment (junior vs regular K) and prediction.

For each test, Table 8 presents means and standard deviations as well as the correlations between screening and MRT2 scores for the quantitative subtest and pre-reading composite. Among the three screening tests, the Daberon and the Brigance were fairly comparable predictors of first-grade readiness as measured by the MRT2. The Daberon accounted for one-fifth of the variation in both the quantitative and pre-reading subtests. The Brigance predicted similarly well on the pre-reading composite, but was a slightly weaker correlate with the quantitative subtest (15% explained variance). The DIALR showed to be the weakest as it explained less than 15% of the variance in both quantitative subtest and pre-reading composite scores.

Table 8 about here

Predicting cognitive development

Scores on the Cognitive Abilities Test (CogAT) were available for 93% of the sample (n=1,194). For each screening test, Table 9 gives the means and standard deviations for the CogAt composite scores. Correlations are reported for the entire group as well as for those who took the regular K - first grade route. The former reveal the extent to which screening test performance predicted cognitive development measured one to two years later. The DIALR correlated most strongly with the CogAT composite, accounting for over one-third of the variance in the latter. The KIDS, Brigance, and Daberon each explained less than one-fourth of the variance in CogAT composite scores.

Table 9 about here

Since the entire-group correlations confound treatment with prediction, the correlation for the K-1 group permits a less prejudiced look at the correspondence between screening test and predict CogAT performance. In theory, the K-1 correlations should be somewhat lower due to the influence of range restriction. As shown in Table 9, however, the correlations did not change appreciably; the DIALR correlation decreased the most (from .60 to .57) and explained variance dropped by only 4%.

Discussion

The preceding analyses shed light on several concerns about tests used for kindergarten entrance. Group membership was related to performance on each of the four screening tests and in some cases, the differences between groups were quite discrepant. Male, minority, low SES, and young examinees were much more likely to be identified as "unready" or "at risk." When these tests are used for placement decisions, one may witness disproportionate placement of these groups in special programs. Indeed, we found that proportionately more males, minorities, low SES and young children are placed in junior kindergartens (Walsh, Ellwein, Eads, & Miller, 1990). We are not so much concerned about labeling these tests as "biased" as we are in pointing out that uncritical acceptance and use of these scores may reinforce if not exacerbate tendencies to create ghetto junior kindergartens.

Although the Brigance, Daberon, and DIALR scores were more consistent indicators than we anticipated, they fell short of standards which recommend reliability coefficients of .90 or higher when making placement decisions (Hopkins & Stanley, 1981). Moreover, evidence on the KIDS warrants caution when interpreting scores for minority examinees and those younger than five years when tested. For these groups, "true" scores could be nearly 20% higher or lower than those obtained. Such inconsistency corrupts

the integrity of placement decisions when they are based, in full or in part, on test performance.

None of the four screening tests were impressive predictors of later measures. Although the Brigance and Daberon were consistent across the CogAT and MRT2, neither contributed to more than one-fifth of the variance in either of the outcome measures. Curiously, performance on the DIALR was barely related to that on the MRT2, but among the four screening tests, it was the best predictor of CogAT scores. Nevertheless, two-thirds of the variance in CogAT scores was left unexplained after considering DIALR performance.

When children are placed on the basis of tests with weak predictive validity, classification errors are a primary concern. Consider a school district which places one-third of the children with the lowest scores into junior kindergarten classrooms. If the test was a robust predictor, we would expect that these children would maintain their relative standing and score in the bottom third of later measures. If the test was a weak predictor, many of those children initially identified as problematic would <u>not</u> score in the lowest third on later measures.

For example, Graue and Shepard (1989) showed that among the lowest third of scorers on the Gesell School Readiness Test (GSRT), only 41% would be in the bottom third of the first-grade class (p. 314). Considering the correlations with end-of-

kindergarten performance (MRT2 pre-reading composite) observed in this study, use of the DIALR would not be much better than the GSRT: only 45% of the children would be identified correctly (Taylor & Russell, 1939). In other words, only four of nine children would score, as predicted, in the lowest third of the MRT2. The risk of misidentification decreases somewhat with use of the Brigance or Daberon: only 50% of the children labeled "at risk" would actually score in the lowest third of the MRT2. The picture does not improve when one considers correlations with early first grade performance (CogAT). Use of the strongest correlate, the DIALR (r=.60), would lead to the misidentification of 42% of the "unready" children. Slight improvements notwithstanding, the predictive validity of the tests examined here are no more accurate than the toss of a fair coin.

Classification errors may be less problematic if junior kindergarten placements result in demonstrable academic, emotional, and social benefits for the child. Moreover, these benefits must be achieved at some reasonable cost to the system. Preliminary analyses of data from the parent study of the one reported here indicate that the benefits of junior kindergarten placements are meager. When compared to their matched peers who entered regular kindergarten, those who took the junior K-regular K route scored only four percentile points higher on the CogAT (Eads, 1989). The estimated cost of this benefit was in excess of $78 million.

Analyses presented here and elsewhere demonstrate that placements based on readiness test performance are not defensible. Readiness tests can provide information for instructional and curriculum planning, but they cannot bear the weight of inflexible placement policies and practices. Although the psychometric properties of these measures may be tinkered with and adjusted, the answer is not one of "building a better mousetrap." Improvements will likely be incremental for the problems in assessing readiness run much deeper. Produced from child-deficit or medical models, these tests are concerned only with identifying and ameliorating an "unready" child. Recent research indicates that readiness is not simply a child characteristic which is amenable to quick and unambiguous measurement (Eisenhart & Graue, in press; Smith & Shepard, 1988). These researchers have shown that readiness may be conceived as a set of shared beliefs constructed by teachers and parents as they interact within particular schools and communities. Placements based on information provided by existing readiness tests discourages sensitivity to these issues and use of such limits our understanding of how some children come to be viewed and treated as not ready.

Table 1

Skill by test matrix

	Brigance	Daberon	DIALR	KIDS
Gross motor skills	*	*	*	*
Fine motor skills	*	*	*	*
Color recognition	*	*	*	*
Number concepts	*	*	*	*
Personal data	*	*	*	*
Visual discrimination	*	*	*	*
Language concepts	*	*	*	*
Classification skills		*	*	*
Able to follow directions	*	*		*
Knowledge of body parts	*	*	*	
Articulation	*		*	
Letter identification			*	*
Auditory discrimination				*

Table 2

Technical information provided in test manuals

	Brigance	Daberon	DIALR	KIDS
Sample characteristics				
Number of states	14	15	6	1
Size of sample	?	1,358	2,447	4,709
Percent white	?	83%	56%	15%
Percent aged 48-60 mos	?	52%	31%	39%
Reliability				
Internal consistency	--	.98	.96	.83[a]
Test-retest	--	.92	.87	--
Validity Studies				
Content Validity?	Yes	Yes	Yes	Yes
Predictive?	No	Yes[b]	No	No

[a] Median of reliability coefficients reported for subtests. No total test reliability coefficients were reported.

[b] Criterion was ratings on a behaviors checklist completed by first grade teachers.

Table 3

Sample district and student characteristics

District characteristics						Examinee characteristics									
										1981 birth quarter					
ID	Size	SES[a]	Kindergarten programs	Test	Cut-score?	Number	Male	Minority	SES[a]	Before	1st	2nd	3rd	4th	Regular K in 86/87
11	1,092	.30	Regular only	DIALR	No	65	52%	0%	37%	11%	28%	22%	22%	19%	100%
50	1,364	.39	Regular only	Brigance	No	82	46%	47%	43%	17%	13%	26%	22%	22%	96%
54	3,499	.41	JrK & regular	Brigance	Yes	277	57%	36%	42%	14%	25%	20%	24%	17%	78%
63	1,852	.30	JrK & regular	Brigance	Yes	131	48%	24%	21%	5%	23%	28%	26%	17%	89%
72	2,099	.21	JrK & regular	KIDS	Yes	143	53%	15%	17%	17%	23%	21%	25%	16%	83%
87	2,533	.51	JrK & regular	Brigance	Yes	195	59%	60%	50%	6%	22%	24%	24%	24%	66%
90	1,204	.47	JrK & regular	Brigance	No	107	58%	67%	50%	10%	19%	19%	33%	20%	71%
99	3,311	.36	JrK & regular	Daberon	Yes	234	45%	8%	38%	15%	23%	21%	25%	16%	64%
207	708	.16	Regular only	DIALR	No	55	51%	17%	11%	18%	18%	29%	27%	7%	100%

[a] The proportion of children eligible for subsidized (free and reduced) lunch in grades K-12.

Table 4

Demographic characteristics of samples by screening test

	Brigance	Daberon	DIALR	KIDS
Male	437 (55%)	134 (46%)	34 (52%)	76 (53%)
Female	355 (45%)	155 (54%)	31 (48%)	67 (47%)
Minority	356 (45%)	26 (10%)	0	22 (15%)
White	433 (55%)	249 (90%)	65 (100%)	121 (85%)
Free lunch	330 (42%)	93 (32%)	24 (37%)	25 (17%)
Not eligible	459 (58%)	194 (68%)	41 (63%)	118 (83%)
Born:				
Before 1981	83 (10%)	45 (16%)	7 (11%)	24 (17%)
Winter, 1981	174 (22%)	63 (22%)	18 (28%)	33 (23%)
Spring, 1981	180 (23%)	65 (22%)	14 (22%)	30 (21%)
Summer, 1981	199 (25%)	74 (26%)	14 (22%)	33 (23%)
Fall, 1981	156 (20%)	42 (15%)	12 (18%)	23 (16%)
1986 placement:				
Not enrolled	3 (--)	0 (--)	0 (--)	1 (--)
Junior K	171 (22%)	84 (29%)	NA	18 (13%)
Regular K	614 (78%)	204 (71%)	65 (100%)	124 (87%)
1987 placement:				
Regular K	111 (15%)	66 (23%)	0	18 (13%)
Repeat K	17 (2%)	6 (2%)	2 (3%)	11 (8%)
T-1	40 (5%)	NA	NA	NA
1st or T-1	587 (77%)	215 (75%)	59 (97%)	106 (77%)
Special Ed	6 (1%)	2 (2%)	0	0

Table 5

Means and standard deviations of screening test scores

	Brigance			Daberon			DIALR			KIDS		
	n	M	(SD)	n	M	(SD)	n	M	(SD)	n	M	(SD)
Total group	788	77	(16)	287	103	(15)	62	82	(9)	143	59	(18)
Male	434	75	(17)	133	102	(14)	32	80	(8)	76	58	(18)
Female	354	80	(14)	154	104	(16)	30	85	(8)	67	61	(17)
Minority	355	72	(17)	26	101	(15)	--	--	-	22	42	(19)
White	430	82	(14)	249	103	(15)	62	82	(9)	121	62	(15)
Free lunch[a]	329	71	(17)	93	93	(19)	22	79	(9)	25	52	(23)
Not eligible	456	82	(15)	192	108	(10)	40	84	(8)	118	61	(16)
Before 1981	82	82	(16)	45	109	(11)	4	88	(4)	24	69	(14)
Winter, 1981	174	82	(14)	63	107	(11)	18	85	(7)	33	67	(13)
Spring, 1981	180	80	(16)	65	105	(13)	14	84	(6)	30	61	(18)
Summer, 1981	196	75	(18)	72	101	(15)	14	79	(10)	33	54	(16)
Fall, 1981	156	69	(17)	42	93	(22)	12	77	(10)	23	42	(12)

[a] Includes any children eligible for lunch subsidy, free or reduced.

Table 6

Demographic effect sizes by screening test[a]

		Effect Size (Δ)			
Group #1	Group #2	Brigance	Daberon	DIALR	KIDS
Males	Females	-0.31	-0.13[b]	-0.55[b]	-0.16[b]
Minorities	Whites	-0.63	-0.13	---	-1.11
Free lunch	Not eligible	-0.69	-1.00	-0.56	-0.50[b]
Before 1981	Fall 1981	+0.81	+1.07	+1.22	+1.50
Winter	Fall	+0.81	+0.93	+1.05	+1.39
Spring	Fall	+0.69	+0.80	+0.78	+1.06
Summer	Fall	+0.38	+0.53	+0.22	+0.67

[a] Where ES = $X_1 - X_2 / SD_{1+2}$

[b] $p > .05$

Table 7

Screening test reliability coefficients and standard errors

	Entire sample	Gender		Ethnicity		Age	
		Male	Female	Black	White	Under 5	Over 5
	Cronbach's alpha coefficient						
Brigance	.81	.82	.79	.81	.81	.80	.81
Daberon	.86	.85	.86	.90	.85	.86	.86
DIALR	.89	.88	.90	na	na	.87	.74
KIDS	.82	.83	.80	.86	.80	.82	.81
	Standard error of measurement (SEM)[a]						
Brigance	8%	8%	8%	8%	7%	8%	7%
Daberon	5%	5%	4%	5%	5%	5%	4%
DIALR	5%	5%	4%	na	na	5%	3%
KIDS	14%	15%	13%	17%	12%	17%	7%

[a] Expressed as percent of total possible items correct.

Table 8

Total screening score as predictor of Metropolitan Readiness Test[a]

		Metropolitan Readiness Subtests									
		Screen		Quantitative				Pre-Reading			
	n	M	SD	M	SD	r^b	EV[c]	M	SD	r^b	EV[c]
Brigance	176	83	11	17	4	.39	15%	57	12	.44	19%
Daberon	192	110	7	18	3	.44	19%	62	8	.46	21%
DIALR	53	82	8	18	4	.36	13%	61	7	.35	12%

[a] Based only on those placed in regular K in 1986/87

[b] $p < .01$

[c] Where $EV = r^2 * 100\%$

Table 9

Total screening score as predictor of Cognitive Abilities Test

		Screen		CcgAT			
	n	M	SD	M	SD	r[a]	EV[b]
Brigance							
Total	737	78	16	186	18	.43	18%
K-1	546	84	12	189	18	.43	18%
Daberon							
Total	270	104	14	193	18	.41	17%
K-1	192	111	6	196	18	.41	17%
DIALR							
Total	58	83	9	194	19	.60	36%
K-1	56	83	8	195	19	.57	32%
KIDS							
Total	129	61	16	193	17	.47	22%
K-1	102	64	14	195	17	.47	22%

[a] $p < .001$

[b] Where $EV = r^2 * 100\%$

References

Brigance, A. H. (1982). Brigance K & 1 Screen for Kindergarten and First Grade. Billerica, MA: Curriculum Associates, Inc.

Butler, S., March, H., Sheppard, M., & Sheppard, J. (1982). Early prediction of reading achievement with the Sheppard School Entry Screening Test. Journal of Educational Psychology, 74, 780-790.

Bradley, R.H. (1985). Review of Gesell School Readiness Test. In J. V. Mitchell, Jr. (Ed.), Ninth mental measurements yearbook (Vol. 1, pp. 609-610). Lincoln, NE: Buros Institute of Mental Measurements.

Cannella, G. S. & Reiff, J. C. (1989). Mandating early childhood entrance/retention assessment: Practices in the 50 states. Paper presented at the 1989 annual meeting of the American Education Research Association, San Francisco.

Cronbach, L. J. (1980b). Validity on parole: How can we go straight? New Directions for Testing and Measurement, 5, 99-108.

Cronbach, L. J. (1980a). Selection theory for a political world. Public Personnel Management, 9(1), 37-50.

Cronbach, L. J. (1984). Essentials of Psychological Testing, 4th edition. New York: Harper and Row.

Danzer, V. A., Gerber, M. F., & Lyons, T. M. (1982). Daberon: A Screening Device for School Readiness. Test manual. Portland, OR: A.S.I.E.P Education Co.

Eads, G. M., Miller, A., Ellwein, M. C., & Walsh, D. J. (1990). Testing for kindergarten readiness: State action and district response. Paper presented at the 1990 annual meeting of the American Education Research Association, Boston.

Eads, G. M. (December, 1989). Research on kindergarten programs by the Department of Education: A report to the legislature. Virginia Department of Education: Division of Research and Testing.

Eisenhart, M. A. & Graue, M. E. (in press). Socially-constructed readiness for school. Qualitative Studies in Education.

Glass, G.V & Hopkins, K.D. (1984). Statistical methods in education and psychology. Second edition. Englewood Cliffs, NJ: Prentice Hall.

Gordon, R. R. (1988). Increasing efficiency and effectiveness in predicting second-grade achievement using a kindergarten screening battery. Journal of Educational Research, 81(4), 238-244.

Graue, M. E. & Shepard, L. A. (1989). Predictive validity of the Gesell School Readiness Tests. Early Childhood Research Quarterly, 4, 303-315.

Gredler, G.R. (1984). Transition class: A viable alternative for the at-risk child? Psychology in the Schools, 21, 463-470.

Harrison, P. & Naglieri, J. (1981). Comparison of the predictive validities of two McCarthy short forms. Psychology in the Schools, 18, 389-393.

Hopkins, K.D. & Stanley, J.C. (1981). Educational and psychological measurement and evaluation. Sixth edition. Englewood Cliffs, NJ: Prentice Hall.

Horn, W. F. & Packard, T. (1985). Early identification of learning problems: A meta-analysis. Journal of Educational Psychology, 77 (5), 597-607.

House, E.R. (1980). Evaluating with validity. Beverly Hills, CA: SAGE.

Kaufman, A.S., & Kaufman, N.L. (1972). Tests built from Piaget's and Gesell's tasks as predictors of first grade achievement. Child Development, 43, 521-535.

Kaufman, N.L. (1985). Review of Gesell School Readiness Test. In J. V. Mitchell, Jr. (Ed.), Ninth mental measurements yearbook (Vol. 1, pp. 607-608). Lincoln, NE: Buros Institute of Mental Measurements.

Loyd, B. (1990) Personal communication.

Mardell-Czudonowski, C.D. & Goldenberg, D.S. (1983). Developmental Indicators for the Assessment of Learning, Revised. Test manual. Edison, NJ: Childcraft Education Corp.

Missouri Department of Elementary and Secondary Education. (1981) Missouri Kindergarten Inventory of Developmental Skills--Alternate Form Guidebook. Jefferson City, MO: Author.

May, D. (1986). Relationships between the Gesell School Readiness Test and standardized achievement and intelligence measures. Educational and Psychological Measurement, 46, 1051-1059.

Meisels, S. J. (1989). High-stakes testing in kindergarten. Educational Leadership, 46, 16-22.

Naglieri, J.A. (1985). Review of Gesell School Readiness Test. In J. V. Mitchell, Jr. (Ed.), Ninth mental measurements yearbook (Vol. 1, pp. 608-609). Lincoln, NE: Buros Institute of Mental Measurements.

National Association for the Education of Young Children. (1988). NAEYC position statement on standardized testing of young children 3 through 8 years of age. Young Children, 43(3), 42-47.

National Association of Early Childhood Specialists in State Departments of Education. (1987) Unacceptable trends in kindergarten entry and placement. Lincoln, NE: Author.

Nurss, J. R. & McGauvran, M. E. (1986). Metropolitan Readiness Tests (Levels 1 and 2) Norms Booklet. The Psychological Corporation: Harcourt Brace Jovanovich.

Popovics, A.J. 91982). Total and component actuarial scores of the Gesell Copy Form Test as predictors of achievement, intelligence, and creativity measures. Journal of Psychology, 110, 293-295.

Rubin, R., Barlow, B., Dorle, J., & Rosen, M. 91978). Preschool prediction of low achievement in basic school skills. Journal of Learning Disabilities, 11, 664-667.

SAS Institute. (1985). SAS User's Guide: Statistics, Version 5. Cary, NC: Author.

Schmidt, S. & Perino, J. (1985). Kindergarten screening results as predictors of academic achievement, potential, and placement in second grade. Psychology in the Schools, 22, 146-151.

Shepard, L. A. & Smith, M. L. (1985). The Boulder Valley kindergarten study: Retention practices and retention effects. Boulder, CO: Boulder Valley Public Schools

Shepard, L.A. & Smith, M.L. (1987). Effects of kindergarten retention at the end of first grade. Psychology in the Schools, 24, 346-357.

Shepard, L. A. & Smith, M. L. (1988). Escalating demand in kindergarten: Counterproductive policies. Elementary School Journal, 89(2), 135-145.

Smith, M. L. & Shepard, L.A. (1988). Kindergarten readiness and retention: A qualitative study of teachers' beliefs and practices. American Educational Research Journal, 25, 307-333.

Stoner, S. & Purcell, K. (1985). The concurrent validity of teachers' judgments of the abilities of preschoolers in a daycare setting. Educational and Psychological Measurement, 45, 421-423.

Taylor, H.C. & Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied Psychology, 23, 565-578.

Thorndike, R. L. & Hagen, E. (1987). Cognitive Abilities Test, Technical Manual, Form 4, Levels 1-2. Chicago: Riverside Publishing.

Walsh, D.J., Ellwein, M.C., Eads, G.M, & Miller, A. (1990) Knocking on kindergarten's door: Who gets in and who is turned away? Manuscript submitted for publication.

Wood, C., Powell, S., & Knight, R.C. 91984). Predicting school readiness: The validity of developmental age. Journal of Learning Disabilities, 17, 8-11.

Appendix

The following tables present results from the ANOVA for each of the four screening tests. Four main effects (gender, ethnicity, SES, and age cohort) and their interactions were assessed using a 2*2*2*5 design was used for three of the four tests. Since no minorities were represented in the DIALR sample, a 2*2*5 was employed. Unequal cell sizes necessitated the use of partial (Type III) sums of squares (SAS Institute, 1985, pp.87-89).

Brigance ANOVA Table

Source of variation	df	Sums of Squares	F	p
Model	39	54,779.19	6.73	.001
Gender (G)	1	2,046.64	9.80	.002
Ethnicity (E)	1	3,010.72	14.42	.001
SES (S)	1	6,759.41	32.37	.001
Age Cohort (A)	4	11,853.70	14.19	.001
G * E	1	0.16	0.00	.978
G * S	1	131.04	0.63	.428
G * A	4	1,186.85	1.42	.225
E * S	1	133.45	0.64	.424
E * A	4	690.77	0.83	.508
S * A	4	335.86	0.40	.807
G * E * S	1	81.08	0.39	.533
S * E * A	4	1,425.41	1.71	.147
G * S * A	4	2,897.30	3.47	.008
E * S * A	4	220.74	0.26	.901
G * E * S * A	4	230.60	0.28	.894
Residual	742	154,924.01		

R^2 = .26

A brief interpretation of the significant interaction among sex, SES, and age cohort: Although older children scored higher than younger, there were two exceptions. Overage boys eligible for free lunch and non-eligible overage girls did not score higher than younger counterparts.

Daberon ANOVA Table

Source of variation	df	Sums of Squares	F	p
Model	33	26,834.88	5.20	.001
Gender (G)	1	104.31	0.67	.415
Ethnicity (E)	1	700.89	4.48	.035
SES (S)	1	1,403.70	8.97	.003
Age Cohort (A)	4	2,974.72	4.75	.001
G * E	1	260.82	1.67	.198
G * S	1	1,961.31	12.54	.001
G * A	4	2,119.74	3.39	.010
E * S	1	990.63	6.33	.013
E * A	4	1,112.86	1.78	.134
S * A	4	744.98	1.19	.316
G * E * S	1	1,590.79	10.17	.002
G * E * A	2	602.21	1.93	.145
G * S * A	4	2,264.59	3.62	.007
E * S * A	3	488.17	1.04	.375
G * E * S * A	0	0.00	.	.
Residual	239	37.380.64		

R^2 = .42

Brief interpretations of the two significant interactions: The first was between sex, ethnicity, and SES. Although free lunch examinees, minority examinees, and male examinees generally scored lower than their respective counterparts, there were a few exceptions. Minority males on free lunch scored higher than any other group of males--closer inspection revealed that the higher mean was prompted by high scores by two students born before 1981. As for the interaction among sex, SES, and age cohort, poor males and females consistently scored lower than their more affluent counterparts across age groups. The interaction seems to have been prompted by a much lower average score for overage girls on free lunch.

DIALR ANOVA Table

Source of variation	df	Sums of Squares	F	p
Model	17	1,805.12	1.79	.061
Gender (G)	1	59.93	1.01	.320
SES (S)	1	290.67	4.90	.032
Age Cohort (A)	4	683.68	2.88	.033
G * S	1	10.04	0.17	.683
G * A	4	337.35	1.42	.242
S * A	4	99.91	0.42	.792
G * S * A	4	5.17	0.04	.957
Residual	44	2.607.93		

R^2 = .41

KIDS ANOVA Table

Source of variation	df	Sums of Squares	F	p
Model	32	22,614.26	3.68	.001
Gender (G)	1	50.82	0.26	.618
Ethnicity (E)	1	1,864.53	0.72	.002
SES (S)	1	288.03	1.50	.223
Age Cohort (A)	4	7,335.83	9.56	.001
G * E	1	21.30	0.11	.740
G * S	1	1.15	0.01	.938
G * A	4	588.90	0.73	.574
E * S	1	32.07	0.17	.683
E * A	4	1,054.77	1.37	.247
S * A	4	1,267.99	1.65	.166
G * E * S	1	416.06	2.17	.144
G * E * A	3	1,287.84	2.24	.088
G * S * A	4	765.97	1.00	.412
E * S * A	2	505.24	1.32	.272
G * E * S * A	0	0.00	.	.
Residual	110	21,100.17		

R^2 = .52

BEST COPY AVAILABLE

Endnotes

1. As used in Virginia and for the purposes of this paper, the term "screening" does not refer to the practice of identifying children in need of further evaluation for special education services. We use the term to denote testing for kindergarten entry and placement.

2. Performance in and across the six specific CogAT subtests is said to reflect the degree to which a child has developed the following skills or competencies: 1) able to comprehend oral English; 2) able to follow directions; 3) able to hold material in short-term memory; 4) possesses effective strategies to obtain information; 5) possesses store of general information and verbal concepts; 6) able to compare and contrast stimuli according to relative size, position, quantity, shape, and time; 7) able to classify, categorize or order familiar objects; and 8) able to use quantitative and spatial relationships and concepts.

3. Age analyses were conducted in a slightly different manner as compared to preceding and following analyses. Here, reliability coefficients were calculated for those younger and older than five years at time of screening.